



DEUTSCHES
PATENT- UND
MARKENAMT

DOSSIER

Offenlegungsschrift

DE 199 62 218 A 1

Int. Cl. 7:
G 10 L 15/24
G 06 K 9/62

(21) Aktenzeichen: 199 62 218.3
(22) Anmeldetag: 22. 12. 1999
(23) Offenlegungstag: 5. 7. 2001

(71) Anmelder:
Siemens AG, 80333 München, DE

(72) Erfinder:
Roderer, Götz, 93051 Regensburg, DE; Gröger, Dirk,
93053 Regensburg, DE

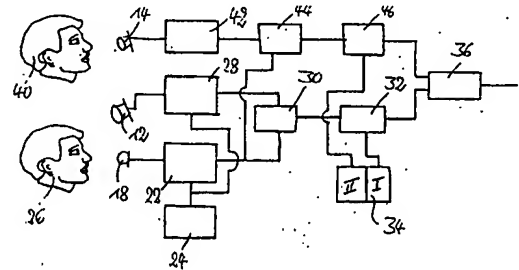
(56) Entgegenhaltungen:
DE 69 101 52 7T2
EP 00 82 304 A1

Die folgenden Angaben sind den vom Anmelder eingereichten Unterlagen entnommen

Prüfungsantrag gem. § 44 PatG ist gestellt

(54) Verfahren und System zum Autorisieren von Sprachbefehlen

(57) In einer mit Sprachbefehlen arbeitenden Einrichtung werden Sprachbefehle dadurch autorisiert, daß vorbestimmten Sprachbefehlen vorbestimmte Orte zugeordnet werden, in denen sich eine den Befehl sprechende Person befinden muß, damit der Sprachbefehl ausgeführt wird. Der Sprachbefehl wird von einem Mikrophon erfaßt. Gleichzeitig wird von einer dem vorbestimmten Ort zugeordneten Kamera die Mundbewegung einer dort sprechenden Person erfaßt und der Befehl zur Ausführung nur freigegeben, wenn die Mundbewegung mit dem vom Mikrophon erfaßten Audiosignal korreliert.



DE 199 62 218 A 1

DE 199 62 218 A 1

Beschreibung

Die Erfindung betrifft ein Verfahren zum Autorisieren von Sprachbefehlen sowie ein System zum Autorisieren von Sprachbefehlen in einer mittels Sprachbefehlen gesteuerten Einrichtung.

Fortschritte auf dem Gebiet der Spracherkennung ermöglichen in zunehmendem Maß eine komfortable Steuerung von Funktionseinheiten über Sprachbefehle. Dazu wird über ein Mikrofon aufgenommene Sprache zunächst im Hinblick auf darin enthaltene typische Laute, Worte oder Wortfolgen analysiert und die festgestellten Laute, Worte oder Wortfolgen werden anschließend mit in einem Befehlsspeicher abgelegten Befehlen verglichen, die in Form von typischen Lauten, Worten oder Wortfolgen abgelegt sind. Bei Übereinstimmung wird der jeweilige Befehl aktiviert.

Aus der EP 0 082 304 A1 ist bekannt, Sprache in Verbindung mit Gesichtserkennung als biometrische Merkmalskombination zu verwenden, um eine Person zu identifizieren und dadurch beispielsweise zu überprüfen, ob eine Person zur Überwindung eines Zugangskontrollsystems autorisiert ist. Dazu werden stimmspezifische Merkmale der zu überprüfenden Person, die ein Schlüsselwort in ein Mikrofon spricht, in einem Korrelationsrechner mit früher gespeicherten stimmspezifischen Merkmalen einer bekannten Person verglichen, die dasselbe Schlüsselwort äußert. Dadurch wird eine erste Korrelationsrate erhalten. Zur gleichen Zeit und zwar wenn ein Schlüsselereignis in der Äußerung des Schlüsselwortes durch die zu überprüfende Person entsteht, wird ein Momentanbild der Mundregion der Person aufgenommen, auf die vorher ein Gittermuster projiziert worden ist. Dieses Momentanbild wird mit einem vorher gespeicherten, entsprechenden Momentanbild der bekannten Person verglichen, um eine zweite Korrelationsrate zu erhalten.

Die beiden Korrelationsraten werden analysiert, um festzustellen, ob die zu überprüfende Person mit der bekannten Person übereinstimmt.

Bei Einrichtungen, die über Sprachbefehle gesteuert werden, insbesondere wenn solche Einrichtungen in Kraftfahrzeugen verwendet werden, besteht ein Problem dahingehend, daß eine andere Person als die Fahrerin oder der Fahrer des Fahrzeugs Sprachbefehle spricht, die dann als Befehle erkannt werden und zur Auslösung bestimmter Funktionen führen. Dies kann zu für den Fahrer verwirrenden oder sogar gefährlichen Zuständen führen.

Der Erfindung liegt die Aufgabe zugrunde, Abhilfe für das vorgenannte Problem zu schaffen.

Gegenstand des Anspruchs 1 ist ein Verfahren zur Lösung dieser Aufgabe.

Erfindungsgemäß wird festgestellt, an welchem Ort sich eine einen Sprachbefehl sprechende Person befindet und werden Sprachbefehle, die diesem Ort zugeordnet sind, nur dann freigegeben, wenn die den jeweiligen Befehl sprechende Person als an dem Ort befindlich erkannt ist. Für die Ortserkennung gibt es unterschiedlichste Möglichkeiten, beispielsweise indem die Sprache von mehreren Mikrofonen erfaßt wird, so daß über ein Triangulationsverfahren der Ort des Sprechers erkannt werden kann.

Ein anderes Verfahren zur Ortserkennung des Sprechers ist im Anspruch 2 angegeben, wobei dieses Verfahren gemäß dem Anspruch 3 zusätzlich dazu verwendet werden kann, die Spracherkennung sicherer zu machen.

Ein System zur Lösung der Erfindungsaufgabe ist im Anspruch 4 angegeben.

Dieses System wird mit den Merkmalen der Ansprüche 5 bis 8 in vorteilhafter Weise weitergebildet.

Die Erfindung ist überall dort mit Vorteil einsetzbar, wo Sprachsteuersysteme eingesetzt werden, bei denen vorbe-

stimmte Befehle nur dann aktiviert werden sollen, wenn sich eine den Befehl sprechende Person innerhalb eines vorbestimmten örtlichen Bereiches befindet.

Die Erfindung kann in Verbindung mit Personenidentifizierungssystemen eingesetzt werden, wie sie beispielsweise in der eingangs genannten EP 0 082 304 A1 beschrieben sind. Auf diese Weise kann zusätzlich überprüft werden, ob die in einem vorbestimmten örtlichen Bereich befindliche Person eine autorisierte Person ist.

Die Erfindung wird im folgenden anhand schematischer Zeichnungen beispielsweise und mit weiteren Einzelheiten erläutert.

Es stellen dar:

Fig. 1 eine Aufsicht auf einen vereinfacht dargestellten Innenraum eines Kraftfahrzeugs, und

Fig. 2 ein Blockschaltbild des erfindungsgemäßen Systems. Gemäß Fig. 1 sind im Innenraum eines Kraftfahrzeugs ein Fahrersitz 2, ein Beifahrersitz 4 und eine Fondsitzbank 6 angeordnet.

Im Dachbereich oberhalb der Windschutzscheibe befindet sich eine Kameraeinrichtung 10 mit Kameras 12 und 14. Der Bildwinkel Ω_1 der Kamera 12 überstreicht einen Bereich, in dem sich normalerweise der Kopf des Fahrers befindet. Der Bildwinkel Ω_2 der Kamera 14 überstreicht einen Bereich, in dem sich normalerweise der Beifahrer befindet.

Im Dachbereich vor dem Fondraum ist eine weitere Kamera 16 angeordnet, deren Bildwinkel Ω_3 den mittleren Bereich des Fondraums überstreicht. Weiter sind Mikrofone 18 und 20 vorgesehen, wobei das Mikrofon 18 dem Fahrer/Beifahrerraum zugeordnet ist und das Mikrofon 20 dem Fondraum zugeordnet ist.

Fig. 2 zeigt ein Blockschaltbild der Anordnung gemäß Fig. 1.

Das Mikrofon 18 ist an eine Sprachanalyseeinrichtung 22 angeschlossen, in der die empfangene Sprache durch Korrelation mit in einem Sprachspeicher 24 abgelegten typischen Sprachlauten und Worten analysiert wird und an deren Ausgang den analysierten Lauten bzw. Worten entsprechende Signale erzeugt werden.

Die Kamera 12, die das Gesicht der Fahrerin oder des Fahrers 26 erfaßt, ist mit einer Bewegungsanalyseeinrichtung 26 verbunden, in der die von der Kamera 12 erfaßten Mundbewegungen im Hinblick auf typische Mundbewegungen analysiert werden und auf ihre Korrelation mit typischen Mundbewegungen zugeordneten, im Sprachspeicher 24 vorher gespeicherten typischen Lauten oder Worten überprüft werden. Bei positiver Korrelation erscheinen am Ausgang der Bewegungsanalyseeinrichtung 26 den typischen Lauten und/oder Worten entsprechende Signale.

Die Ausgangssignale der Sprachanalyseeinrichtung 22 und der Bewegungsanalyseeinrichtung 26 werden einer Synchronprüfeinrichtung 30 zugeführt, die nur solche Laute und/oder Worte durchläßt und einer Korrelationseinrichtung 32 zuführt, die zeitgleich aus der Sprachanalyseeinrichtung 22 und der Bewegungsanalyseeinrichtung 26 ausgegeben werden.

In der Korrelationseinrichtung 32 werden die empfangenen typischen Laute und/oder Worte mit in einer Sektion I eines Befehlsspeichers 34 abgelegten Befehlen verglichen. Bei positivem Vergleich wird am Ausgang der Korrelationseinrichtung 32 ein entsprechendes Befehlssignal erzeugt und in einer Steuereinrichtung 36 in ein entsprechendes Steuersignal zum Steuern einer Funktion umgewandelt.

Die dem Beifahrer 4 zugeordnete Kamera 14 ist mit einer weiteren Bewegungsanalyseeinrichtung 42 verbunden, der eine Synchronprüfeinrichtung 44 nachgeschaltet ist, die mit der Sprachanalyseeinrichtung 22 und einer weiteren Korrelationseinrichtung 46 verbunden ist, die mit einer Sek-

tion II des Befehlsspeichers 34 und mit der Steuereinrichtung 36 verbunden ist.

Die Kamera 16 des Fondraums ist der Einfachheit halber nicht in das Blockschalbild eingezeichnet. Sie ist ähnlich geschaltet wie die Kamera 14 bzw. die Kamera 12. Das Fondraummikrofon 20 kann unmittelbar an die Sprachanalyseeinrichtung 22 angeschlossen sein.

Die Funktionen der einzelnen entsprechend fortschrittlichen gängigen Verfahren der Spracherkennung, der Bildverarbeitung und der Informationskorrelationen arbeitenden Komponenten sowie deren Aufbau sind an sich bekannt und werden daher nicht erläutert.

Im folgenden wird daher lediglich das erfindungsgemäße Zusammenwirken der Komponenten beschrieben:

Es sei angenommen, der Fahrer 26 spräche das Befehlswort "Scheibenwischer an". Das Sprechen dieses Sprachbefehls wird zum einen vom Mikrofon 18 erfaßt und zum anderen von der Kamera 12. In der Sprachanalyseeinrichtung 22 wird durch Vergleich mit im Sprachspeicher 24 abgelegter Information ermittelt, daß die Wortgruppe "Scheibenwischer an" gesprochen wurde und am Ausgang der Sprachanalyseeinrichtung 22 erscheint ein entsprechendes Signal. In der Bewegungsanalyseeinrichtung 26 werden die Mundbewegungen analysiert, die dem Sprechen der Wortfolge "Scheibenwischer an" entsprechen und durch Korrelation mit im Sprachspeicher 24 abgelegten Informationen festgestellt, daß diese Wortfolge gesprochen wurde, so daß am Ausgang der Bewegungsanalyseeinrichtung 26 ein entsprechendes Signal erzeugt wird. In der Synchronprüfeinrichtung 30 wird überprüft, ob beide Signale synchron zueinander sind, so daß das synchrone Signal zu der Korrelationseinrichtung 32 durchgelassen wird, in der überprüft wird, ob dem Signal ein im Befehlsspeicher 34 abgelegter Befehl entspricht. Ist im Befehlsspeicher 34 ein entsprechender Befehl "Scheibenwischer an" abgelegt, so sendet die Korrelationseinrichtung 32 ein entsprechendes Signal an die Steuereinrichtung 26, in der ein Steuersignal zum Inbetriebsetzen des Scheibenwischers erzeugt wird.

Wird die Wortfolge "Scheibenwischer an" vom Beifahrer 40 gesprochen, so läuft der geschilderte Vorgang in der Sprachanalyseeinrichtung 22, der Bewegungsanalyseeinrichtung 42, der Synchronprüfeinrichtung 44 und der Korrelationseinrichtung 46 ab. In der Korrelationseinrichtung 46 erfolgt keine positive Korrelation mit einem in der Sektion II des Befehlsspeichers 34 abgelegten Befehl, da in der Sektion II nur solche Befehle abgelegt sind, die für den Beifahrer erlaubte Befehle sind. Entsprechend wird die Steuereinrichtung 36 nicht aktiv, wenn der Beifahrer 40 den Befehl "Scheibenwischer an" spricht.

Durch Segmentierung der in dem Befehlsspeicher 34 abgelegten Befehle in einzelne Sektionen, die unterschiedlichen Kameras bzw. örtlichen Bereichen im Fahrzeuginneren zugeordnet sind, wird somit ermöglicht, daß Sprachbefehle nur dann ausgeführt werden, wenn sie aus einem, dem jeweiligen Sprachbefehl zugeordneten räumlichen Bereich des Fahrzeuginnenraums kommen. Dabei können bestimmte Befehle durchaus allen Bereichen des Fahrzeuginnenraums zugeordnet sein beispielsweise "Innenbeleuchtung an" oder "Radio aus" und so weiter. Weiter ist es möglich, daß ein Befehl "Temperatur höher" unterschiedliche Funktionen auslöst, je nachdem, ob er vom Fahrer, Beifahrer oder aus dem Fond gesprochen wird. In einer Heizungs-/Klimaanlage werden dann entsprechende Funktionen ausgelöst, die den örtlichen Bereichen zugeordnet sind.

Es versteht sich weiter, daß der Fahrer beispielsweise mittels einer nicht dargestellten Bedieneinheit Befehle einzelnen örtlichen Bereichen zuordnen kann; beispielsweise kann er Telefonfunktionen für den gesamten Innenraum freige-

ben, nur für den Beifahrer freigeben usw.

In der bzw. den Synchronprüfeinrichtungen 30, 44 können verstümmelte Worte, die bezüglich ihrer Synchronisierung noch erkennbar sind, die aber nicht vollständig aufgenommen sind, beispielsweise weil der Mund der sprechenden Person teilweise verdeckt wird oder ein Geräusch das gesprochene Wort überdeckt, durch gegenseitige Korrelation ergänzt werden, so daß die Qualität der der Korrelationseinrichtung 32 zugeordneten Information bzw. Worte verbessert ist.

Die Erfindung kann in vielfältiger Weise abgeändert werden. Beispielsweise können die den unterschiedlichen Kameras zugeordneten Sprachanalyseeinrichtungen, Synchronprüfeinrichtungen und Korrelationseinrichtungen in einem gemeinsamen Hardware-Baustein zusammengefaßt sein, wobei die von der jeweiligen Kamera erfaßte optische Information hinsichtlich ihrer Kamerazugeordnung gekennzeichnet wird. Die verschiedenen Sektionen des Befehlsspeichers 34 können dadurch ausgebildet sein, daß den insgesamt im Befehlsspeicher 34 befindlichen Befehlen Informationen hinsichtlich der jeweiligen örtlichen Bereiche bzw. Kameras zugefügt sind. Die Synchronprüfung zwischen optischen und akustischen Informationen kann bereits vor der Analyse hinsichtlich vorbestimmter Sprachbausteine erfolgen, da zwischen der gesprochenen Sprache und den erfaßten Mundbewegungen Grundkorrelationen bestehen beispielsweise zwischen Sprachbeginn und Beginn einer Mundbewegung. Der Innenraum des Fahrzeugs kann in mehrere Bereiche aufgeteilt sein, beispielsweise der Fondraum in drei nebeneinander angeordnete Bereiche. Das System kann auf den Außenraum erweitert werden, der von einer oder mehreren Kameras bereichsweise überdeckt wird.

In vereinfachter Ausführungsform kann das System dazu verwendet werden festzustellen, ob eine Sprache überhaupt von einer im Sichtfeld einer Kamera befindlichen Person ausgeht, indem eine Synchronizität zwischen akustisch aufgenommener Sprache und optisch erfaßter Mundbewegung festgestellt oder nur festgestellt wird, ob sich ein Mund bewegt. Liegt keine Mundbewegung vor, so wird eine Sprachauswertung gar nicht aktiviert, so daß Rechnerleistung eingespart bzw. nicht unnötig verbraucht wird.

Patentansprüche

1. Verfahren zum Autorisieren von Sprachbefehlen, bei welchem Verfahren vorbestimmten Sprachbefehlen wenigstens ein vorbestimmter Ort zugeordnet wird, an dem sich eine die Befehle sprechende Person befinden muß, damit die Sprachbefehle ausgeführt werden, festgestellt wird, ob sich eine einen Sprachbefehl sprechende Person an dem Ort befindet und bei positiver Feststellung die Ausführung der Sprachbefehle freigegeben wird.
2. Verfahren nach Anspruch 1, wobei Sprache akustisch aufgenommen und hinsichtlich typischer Laute analysiert wird, wenigstens eine Kamera Mundbewegungen einer an einem vorbestimmten Ort befindlichen Person erfaßt, die aufgenommenen Mundbewegungen hinsichtlich typischer Mundbewegungen analysiert werden, die typischen Mundbewegungen und die typischen Laute hinsichtlich ihrer zeitlichen Übereinstimmung verglichen werden und bei positivem Vergleich der Ort des Sprechers als der Ort der mit der Kamera aufgenommenen Person identifiziert wird.
3. Verfahren nach Anspruch 2, wobei typischen Mundbewegungen typische Laute zugeordnet sind und aku-

stisch schlecht aufgenommene Laute durch typische Laute ersetzt werden, die zeitgleich aufgenommenen, typischen Mundbewegungen zugeordnet sind.

4. System zum Autorisieren von Sprachbefehlen in einer mittels Sprachbefehlen gesteuerten Einrichtung, 5
enthaltend ein Mikrofon (18, 20) zur Aufnahme von Sprache,

eine dem Mikrofon nachgeschaltete Sprachanalyseeinrichtung (22) zur Analyse der Sprache hinsichtlich typischer Laute und/oder Worte, 10

eine Kameraeinrichtung (12, 14, 16), die Mundbewegungen einer an einem vorbestimmten Ort befindlichen Person aufnimmt,

eine der Kamera nachgeschalteten Bewegungsanalyseeinrichtung (28, 42), die die aufgenommenen Mundbewegungen hinsichtlich typischer Mundbewegungen und/oder diesen zugeordneten typischen Lauten und/oder Worten analysiert, 15

eine Synchronprüfeinrichtung (30, 44), die die typischen Mundbewegungen und/oder die aus diesen abgeleiteten typischen Laute und/oder Worte hinsichtlich ihrer zeitlichen Übereinstimmung mit aus der Sprachanalyse hergeleiteten typischen Lauten und/oder Worten vergleicht und bei Übereinstimmung weiterleitet, 20
und eine Korrelationseinrichtung (32, 46), die durch Vergleich der weitergeleiteten typischen Laute und/oder Worte mit in einem Befehlsspeicher (34) gespeicherten Befehlen Befehlssignale an eine Steuereinrichtung (36) weitergibt, die die Befehle in diesen zugeordnete Steuersignale umsetzt. 25 30

5. System nach Anspruch 4, wobei die Kameraeinrichtung (12, 14, 16) derart ausgebildet ist, daß die Mundbewegungen von an unterschiedlichen vorbestimmten Orten befindlichen Orten erfassbar sind und den unterschiedlichen Orten vorbestimmte zulässige Befehle zugeordnet sind. 35

6. System nach Anspruch 4 oder 5, wobei eine Speichereinrichtung (24) vorgesehen ist, in der typischen Mundbewegungen zugeordnete typische Laute und/oder Worte abgelegt sind, und die Korrelationseinrichtung (32, 46) bei schlechter Qualität der akustisch aufgenommenen typischen Laute und/oder Worte die den zeitgleich aufgenommenen Mundbewegungen zugeordneten typischen Laute und/oder Worte weiterleitet. 40

7. System nach einem der Ansprüche 4 bis 6, wobei die Kamera (12, 14, 16) den Fahrer- und Beifahrerraum in einem Kraftfahrzeug erfaßt und die Steuersignale zur Steuerung von Funktionen vorgesehen sind. 45

8. System nach Anspruch 7, wobei eine Kamera (16) den Fondraum des Kraftfahrzeugs erfaßt. 50

Hierzu 2 Seite(n) Zeichnungen

55

60

65

